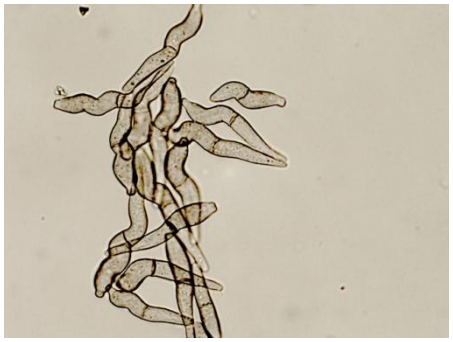
Deciphering the mechanism of emergence of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation

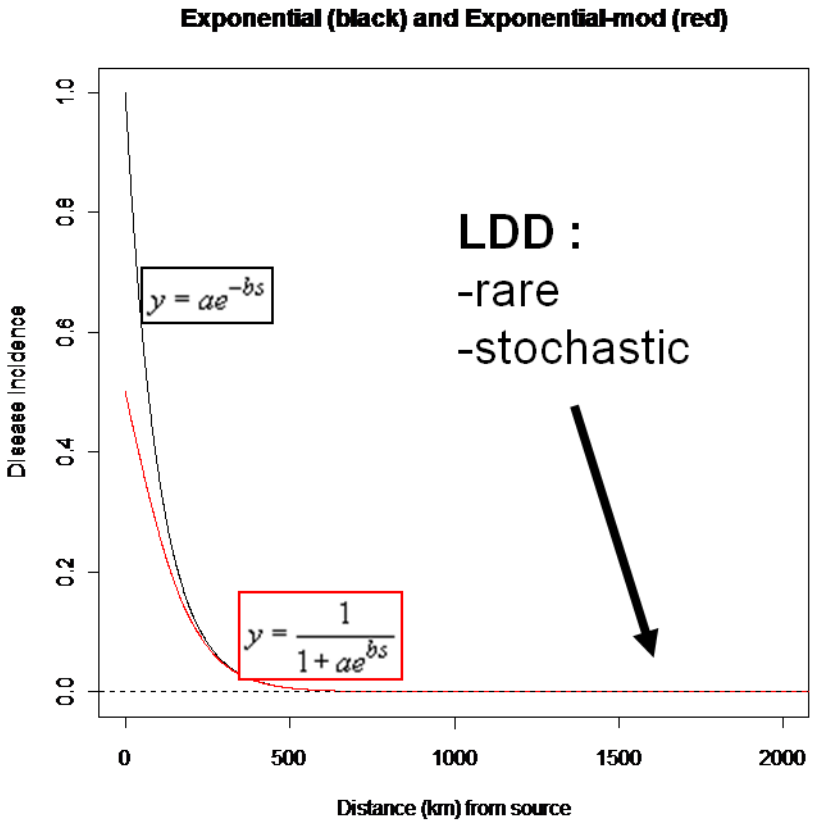
Benoît Barrès, Jean Carlier, Christian Cilas, Marc Seguin, Catherine Fenouillet and Virginie Ravigné



UPR 31 **MICHELIN** **UMR BGPI**

Jacques Monod Conference "New and Emerging Fungal Diseases of Animals and Plants: evolutionary aspects in the context of global changes"

Emerging pathogen and Long Distance Dispersal (LDD)



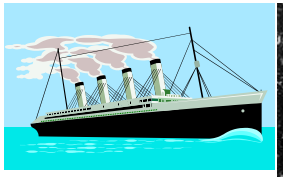
LDD :
-rare
-stochastic

Natural dispersal :

(source : R. C. Kemerait, Jr.)

Phakopsora pachyrhizi, Pan et al. 2006

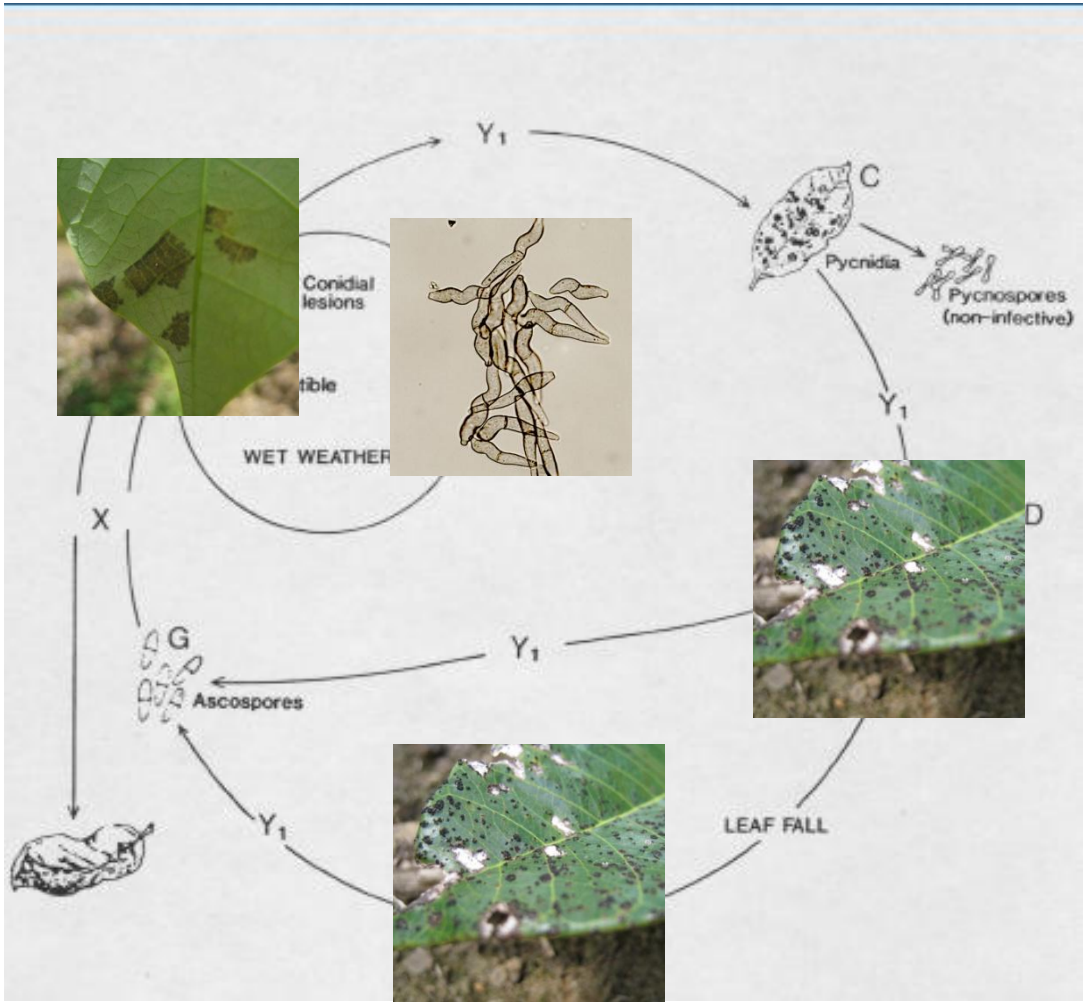
Human mediated dispersal :



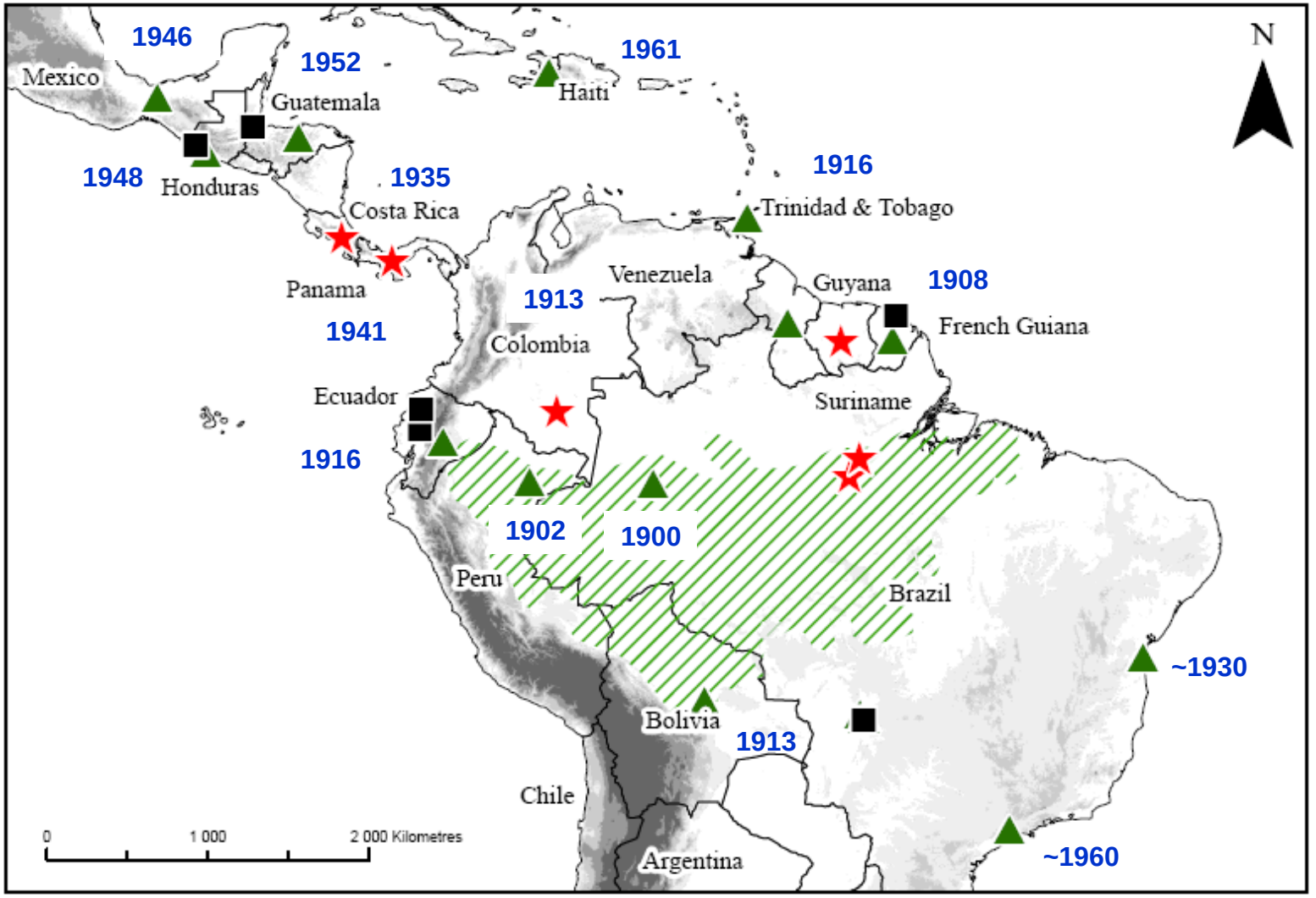
(source : Milgroom lab, chesnut blight commission)

Cryphonectria parasitica, Milgroom et al. 1996

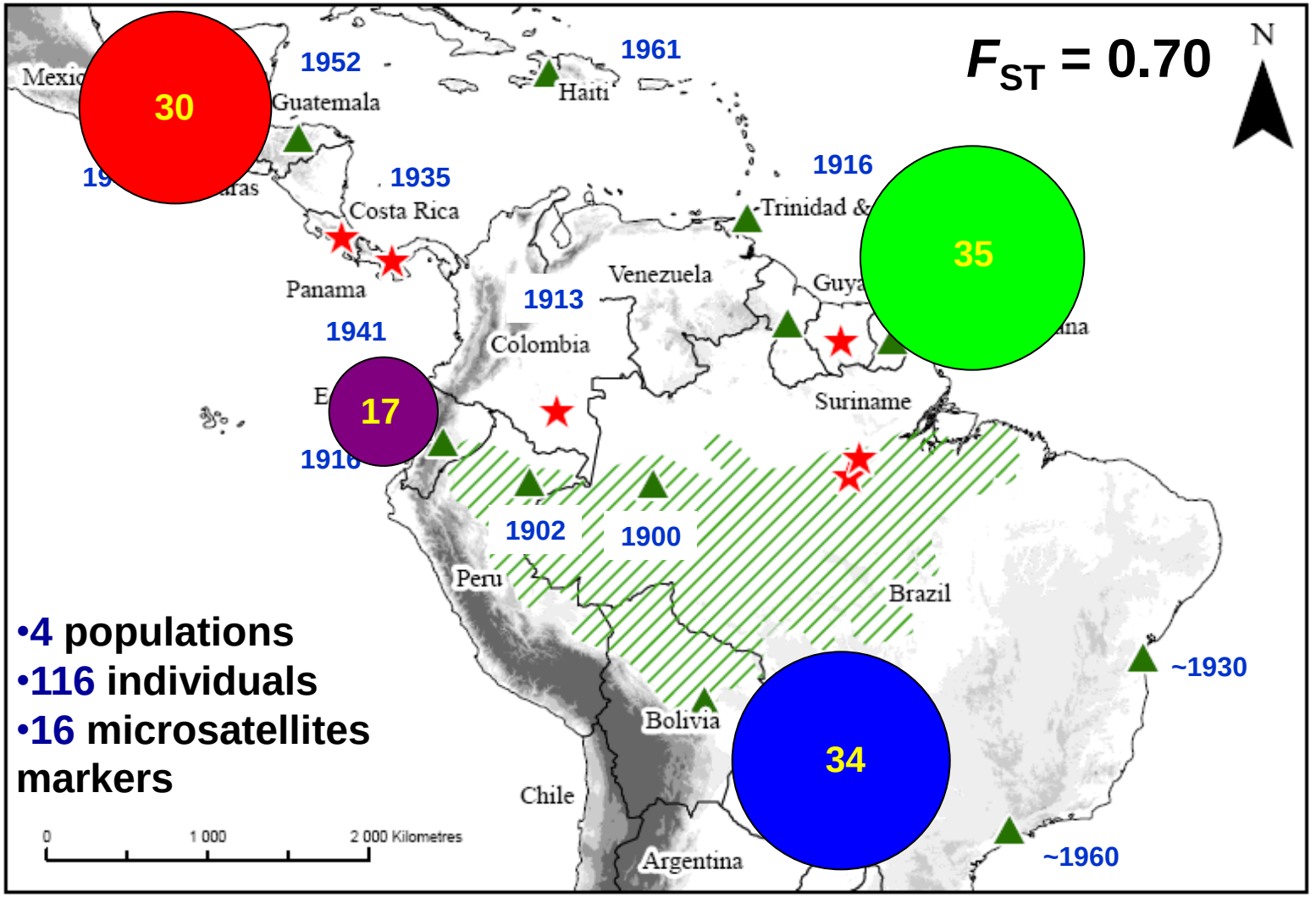
Biological cycle and symptom



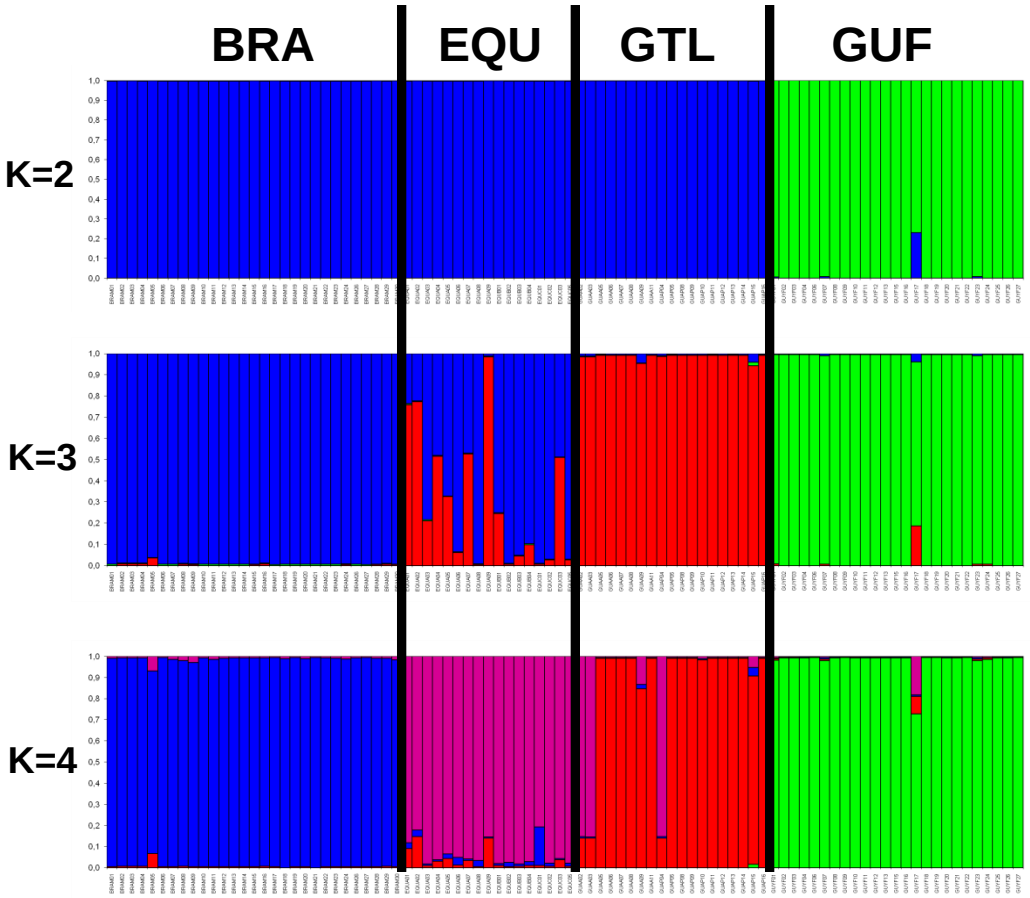
Continental scale dispersion of *M. ulei*



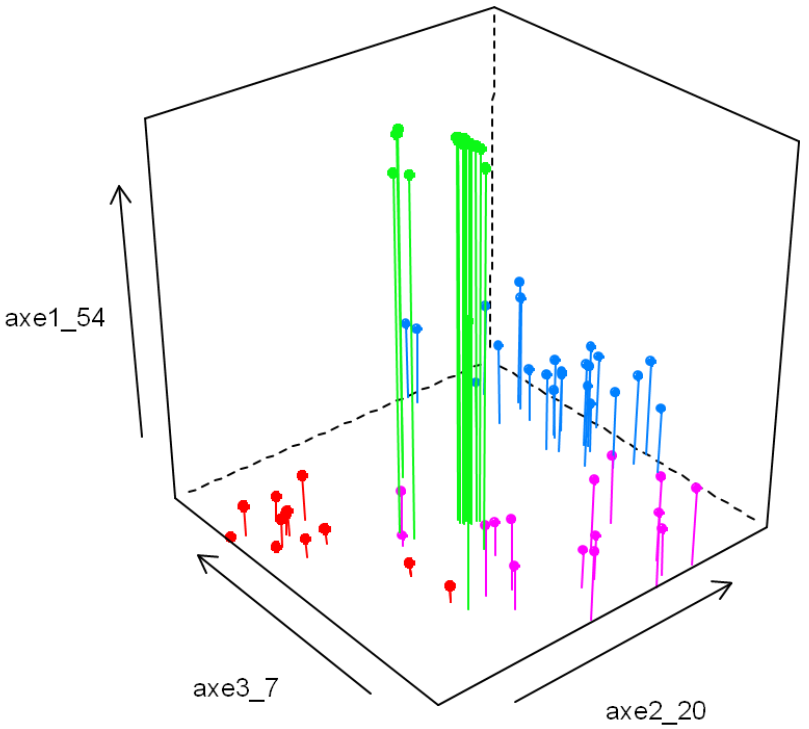
Sampling design



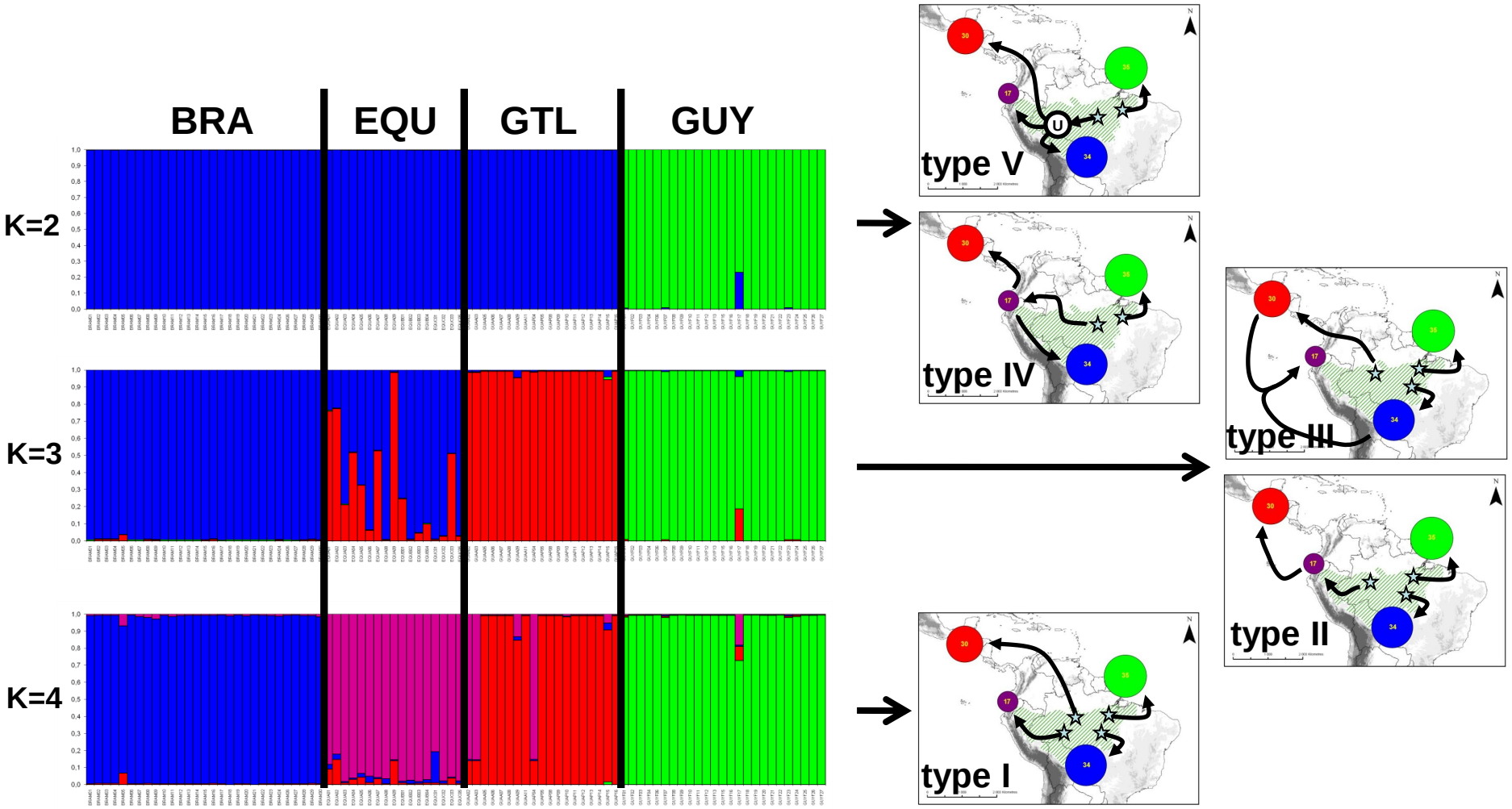
Genetic structure



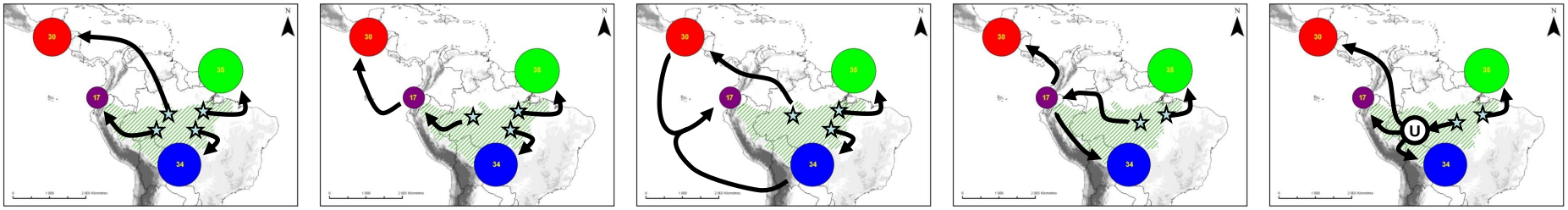
PCA

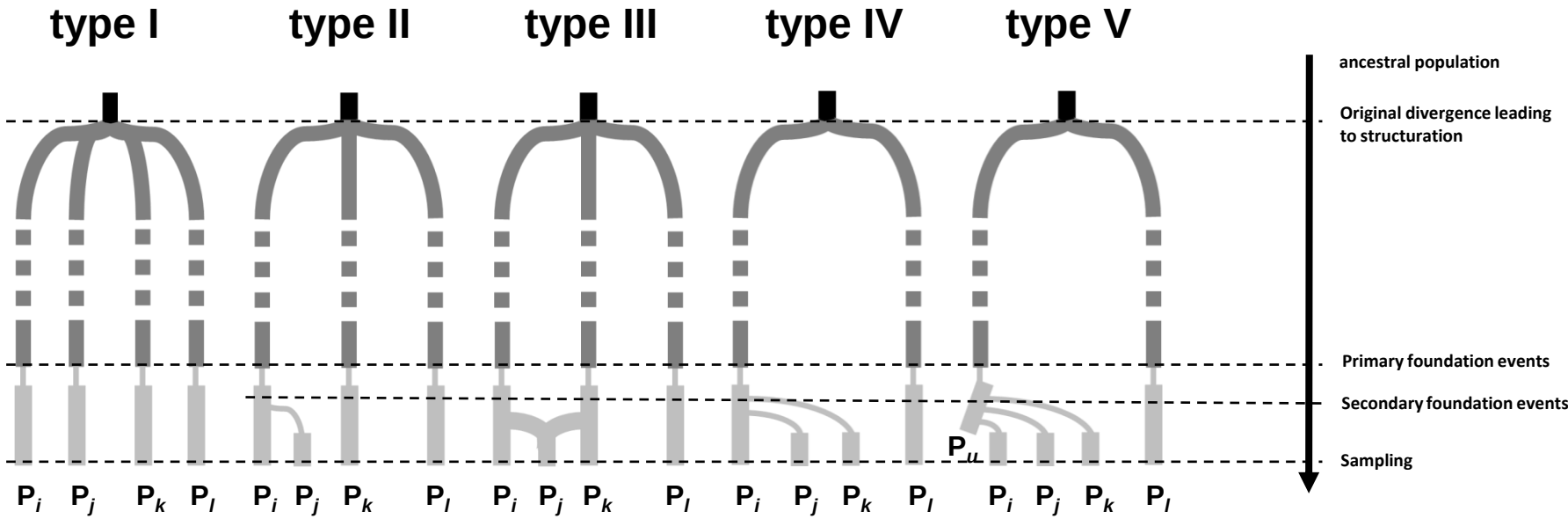


Definition of scenarios

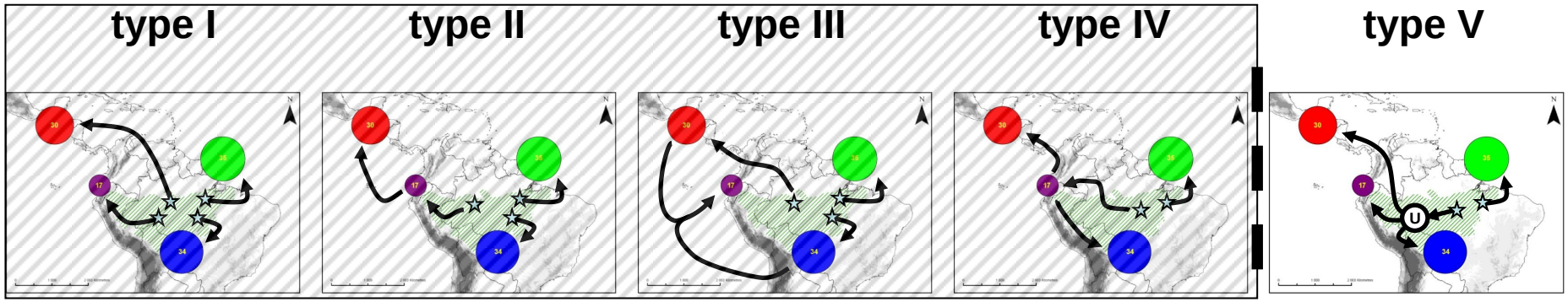


Definition of scenarios



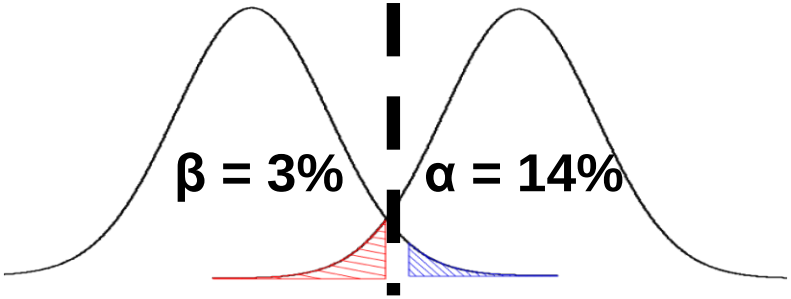


The choice of a scenario and consistency of it



***P*=0.023**
[0.018-0.028]

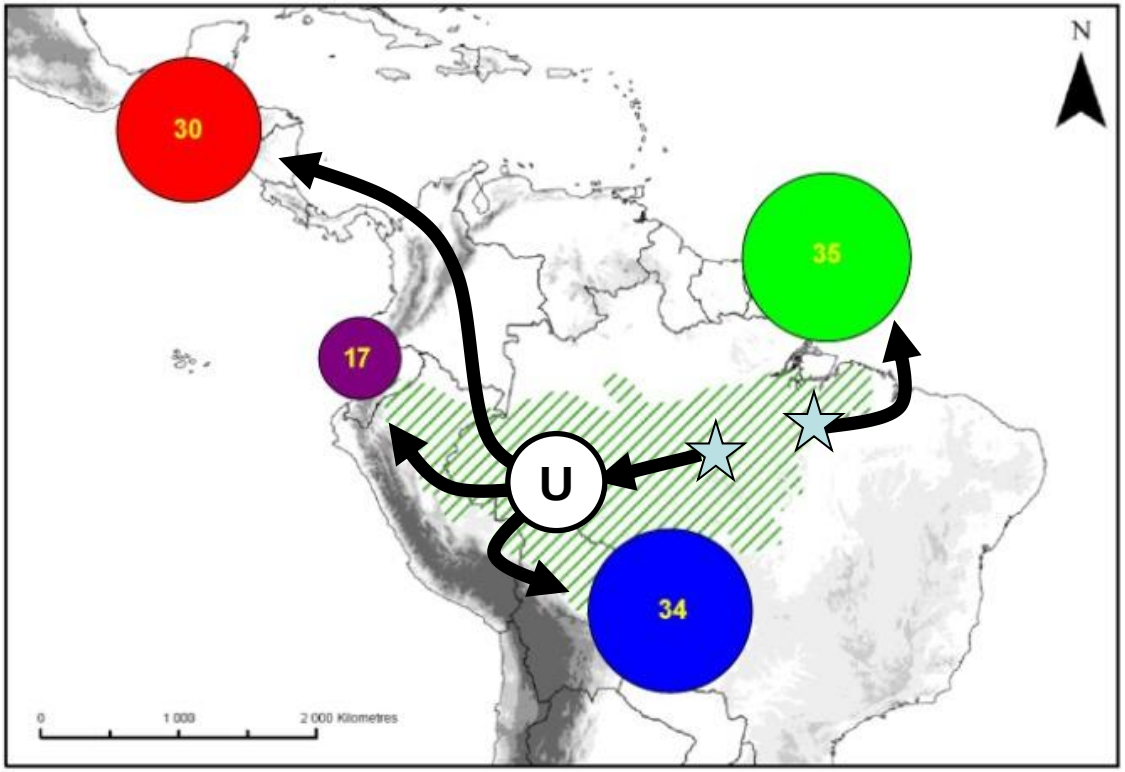
***P*=0.977**
[0.972-0.982]



-Goodness of fit evaluated by Model-checking

-Confidence in the choice of scenario evaluated with pseudo-observed datasets

History of colonization of *M. ulei* in Latin America



-No regular gene flow at the continental scale

-Two independent founder event

-West-populations share a common “ancestry”

Thank you for your attention

UMR BGPI

Jean Carlier
Virginie Ravigné
Elisabeth Fournier
Didier Tharreau
Daniel Bieysse

UPR 31

Catherine Fenouillet
Christian Cilas
Jean Guyot
Franck Rivano
Christophe Couturier

UMR TETIS

Danny Lo Seen
Gérard Lainé

UMR DAP

Marc Seguin
Vincent Le Guen

MICHELIN

Eric Cavaloc
Carlos Mattos
Milton Luiz

Funding : CIRAD, MICHELIN, ANR EMERFUNDIS, ANR EMILE

The ABC framework

3 steps :

1) generating simulated data sets

2) selecting simulated data sets closest to observed data set

3) estimating posterior distributions of parameters through a local linear regression procedure

(*Excoffier* et al., 2005)

Prior distribution of parameters

	Parameter name	distribution	Extremum values
Population effective sizes of the			
ancestral population	N	uniform	{10 – 100,000}
population leading to Brazilian population foundation	NOb	uniform	{10 – 100,000}
population leading to Equatorian population foundation	NOe	uniform	{10 – 100,000}
population leading to Guatemalan population foundation	NOgt	uniform	{10 – 100,000}
population leading to French Guyanian population foundation	NOgy	uniform	{10 – 100,000}
population leading to unsampled population foundation	NOu	uniform	{10 – 100,000}
bottleneck during foundation of Brazilian population	NBb	log-uniform	{1 - 100}
bottleneck during foundation of Equatorian population	NBe	log-uniform	{1 - 100}
bottleneck during foundation of Guatemalan population	NBgt	log-uniform	{1 - 100}
bottleneck during foundation of French Guyana population	NBgy	log-uniform	{1 - 100}
bottleneck during foundation of unsampled population	NBu	log-uniform	{1 - 100}
Brazilian population	Nb	uniform	{10 – 100,000}
Equatorian population	Ne	uniform	{10 – 100,000}
Guatemalan population	Ngt	uniform	{10 – 100,000}
French Guyana population	Ngy	uniform	{10 – 100,000}
Unsampled population	Nu	uniform	{10 – 100,000}
Time of the			
original divergence leading to genetic structure in source population	TO	uniform	{600 – 100,000}
foundation of the Brazilian population	TFb	log-uniform	{16 - 500}
foundation of the Equatorian population	TFe	log-uniform	{16 - 500}
foundation of the Guatemalan population	TFgt	log-uniform	{16 - 500}
foundation of the French Guyana population	TFgy	log-uniform	{16 - 500}
foundation of the unsampled population	TFu	log-uniform	{16 - 500}
secondary foudation of Brazilian population	TSb	log-uniform	{16 - 500}
secondary foudation of Equatorian population	TSe	log-uniform	{16 - 500}
secondary foudation of Guatemalan population	TSgt	log-uniform	{16 - 500}
secondary foudation of French Guyana population	TSgy	log-uniform	{16 - 500}
admixture event	TA	log-uniform	{16 - 500}

Population effective size :

-population UNIF {10-100,000}
-bottleneck LOG-U {1-100}

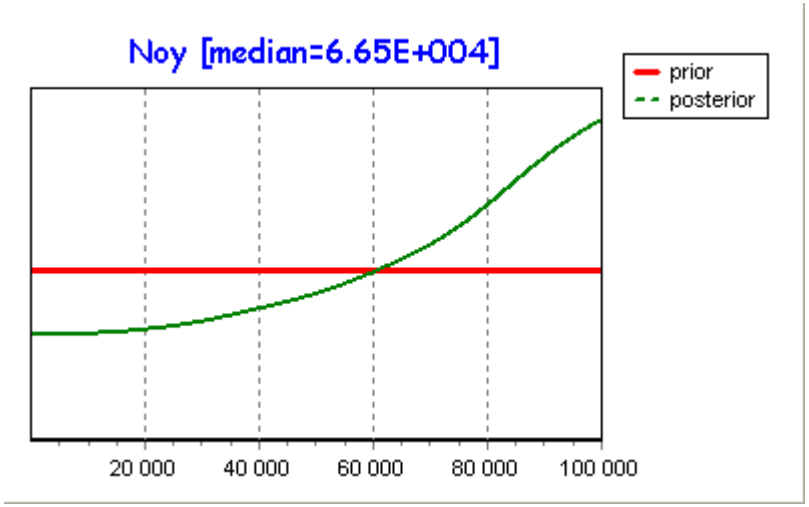
Time of divergence :

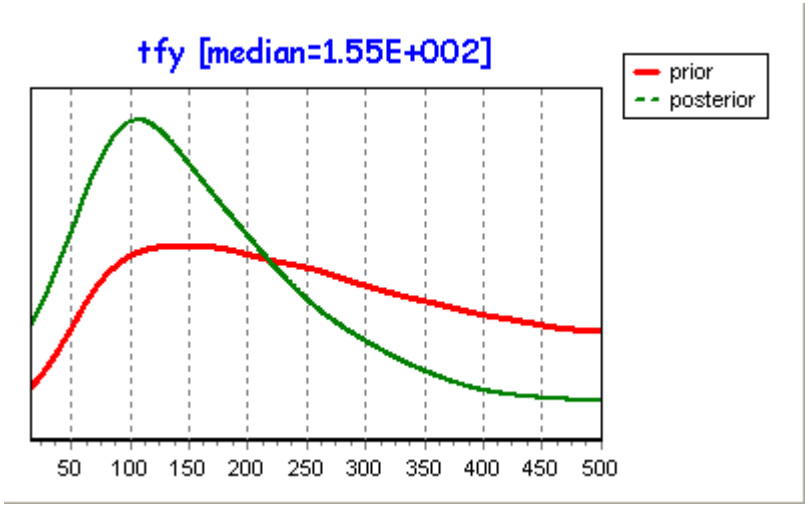
-ancestral UNIF {600-100,000}
-foundation pop LOG-U {16-500}

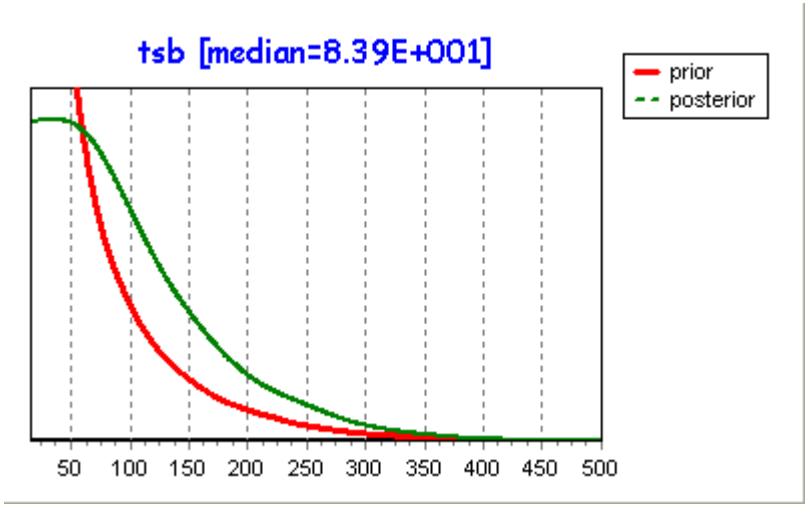
Admixture rate :

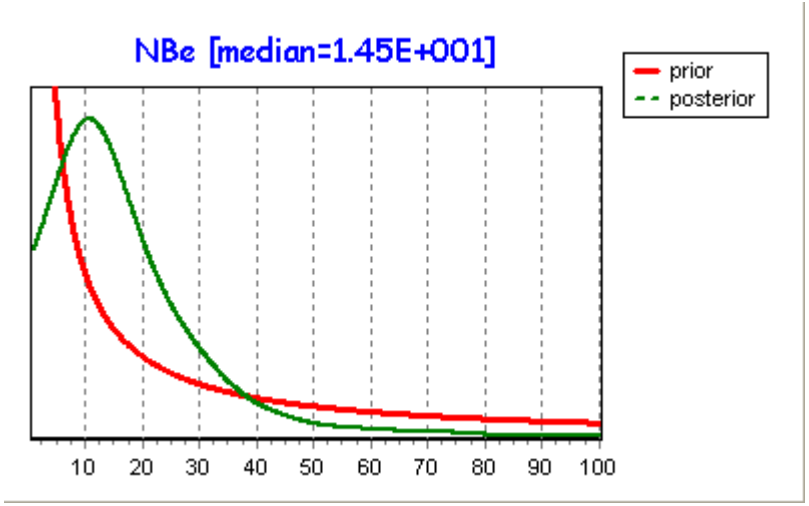
- UNIF {0.001-0.999}

Parameters estimation









Goodness of fit of the model

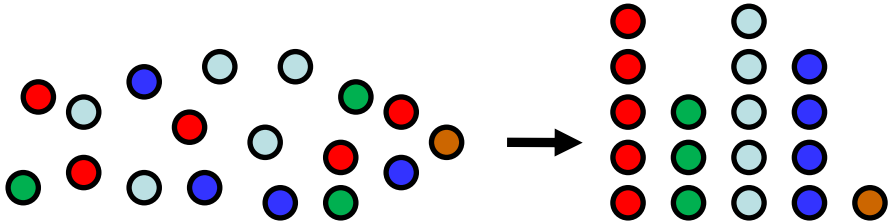
-Model checking

Summary statistics	Observed value	Probability				
		Best scenario of Class I	Best scenario of Class II	Best scenario of Class III	Best scenario of Class IV	Best scenario of Class V
NAL BRA	2.1250	0.6400	0.5855	0.8690	0.9355	0.7815
NAL EQU	1.6250	0.1360	0.1445	**0.0045 ****	0.1805	0.2225
NAL GTM	1.5625	0.4805	0.7230	0.8740	0.8745	0.7060
NAL GUF	2.0625	0.8900	0.9300	0.9420	0.8055	0.6785
HET BRA	0.2336	0.5260	0.3940	0.7115	0.8610	0.5965
HET EQU	0.2775	0.3085	0.2690	**0.0145 ***	0.5105	0.3940
HET GTM	0.1598	0.4675	0.6220	0.8380	0.8120	0.6455
HET GUF	0.2195	0.7225	0.8300	0.8585	0.8060	0.6360
MGW BRA	0.6766	0.1780	0.3270	0.2635	0.2265	0.2970
MGW EQU	1.4648	**0.9820 ***	**0.9660 ***	**1.0000 *****	**0.9525 ***	0.9480
MGW GTM	1.6393	0.9060	0.7150	0.6740	0.7420	0.7930
MGW GUF	0.3677	**0.0040 ****	**0.0040 ****	**0.0090 ****	**0.0100 ***	**0.0190 ***
N2P BRA&EQU	2.6875	0.1565	0.0760	0.5645	0.7910	0.5410
N2P BRA&GTM	2.6875	0.3295	0.2980	0.5990	**0.9625 ***	0.7585
N2P BRA&GUF	3.8125	0.9010	0.8550	**0.9540 ***	0.8940	0.7405
N2P EQU&GTM	2.0000	**0.0105 ***	0.3150	0.0785	0.3910	0.2340
N2P EQU&GUF	3.2500	0.6530	0.5290	0.3685	0.3290	0.3165
N2P GTM&GUF	3.2500	0.8430	0.8955	0.9455	0.7955	0.6445
V2P BRA&EQU	1.2370	0.1645	0.0530	0.0650	0.3790	0.2650
V2P BRA&GTM	1.1814	0.1680	0.0570	**0.0410 ***	0.4250	0.3090
V2P BRA&GUF	27.6895	**0.9970 ****	**0.9910 ****	**0.9800 ***	0.8040	0.8670
V2P EQU&GTM	0.2401	**0.0030 ****	0.0680	**0.0030 ****	0.1325	0.0540
V2P EQU&GUF	24.9809	**0.9980 ****	**0.9870 ***	**0.9660 ***	0.7300	0.8360
V2P GTM&GUF	26.1170	**0.9970 ****	**0.9880 ***	**0.9800 ***	0.7740	0.8520
DAS BRA&EQU	0.5418	**0.9795 ***	**0.9960 ****	0.8385	0.1025	0.5850
DAS BRA&GTM	0.4068	0.8635	**0.9630 ***	**0.9680 ***	**0.0415 ***	0.2510
DAS BRA&GUF	0.1786	0.2345	0.4715	0.5890	0.9100	0.8320
DAS EQU&GTM	0.5856	**0.9910 ****	0.4395	0.4650	0.1435	0.6600
DAS EQU&GUF	0.1583	0.1545	0.4120	0.5090	0.8785	0.8035
DAS GTM&GUF	0.0865	**0.0320 ***	0.1725	0.2260	0.6090	0.5205

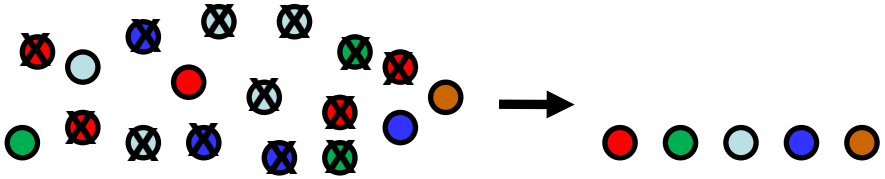
Clone correction

3 datasets :

1) Complete dataset



2) One individual for each multilocus genotype



3) Over-represented multilocus genotype removed